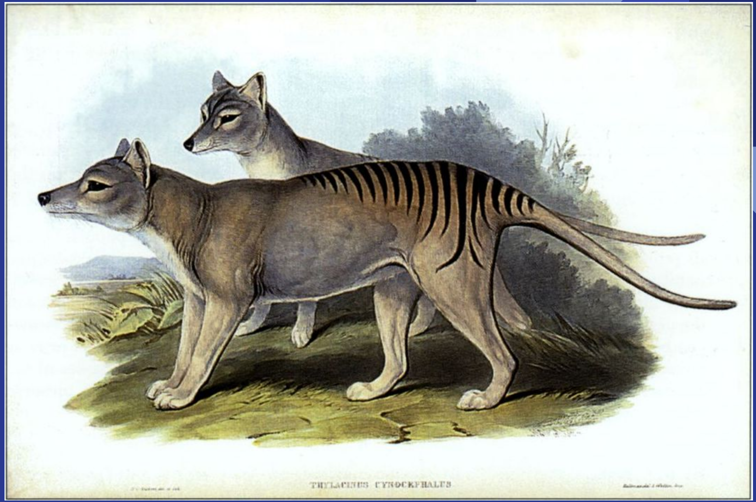
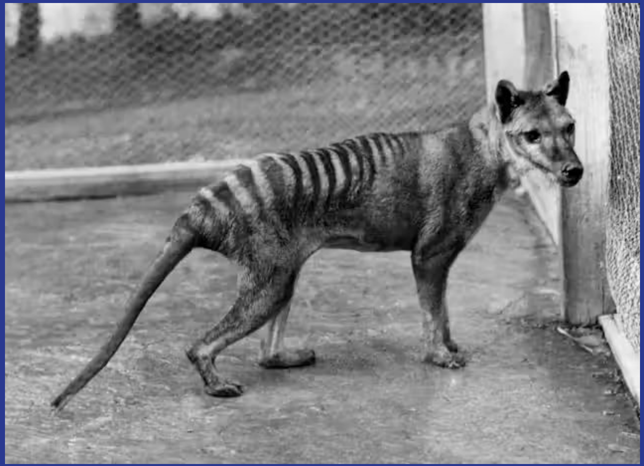
Modeling De-extinction: A Computational Approach to Back-breeding Dynamics

Nimay Gupta

Introduction

- Overview
- The Problem
- Methods Used So Far
- Biological Background of Back Breeding





Overview

Key Question: Can extinct species be brought back using computational models?

De-extinction Focus:

- **De-extinction:** Reviving species using genetic techniques.
- One approach: **Back-breeding,** selectively breeding closely related species to restore genetic traits of extinct species.

Motivation:

- De-extinction is of high ecological and conservation interest. Computational models help estimate feasibility and timelines.

The Problem

- Species are currently going extinct at **100 to 1,000 times the natural rate** due to human impact.
- With many species playing important roles in ecosystems, extinction creates several challenges; it leads to a loss in **biodiversity, diminishes genetic variability, and curbs enthusiasm for conservation projects.**
- Hence, with the emergence of advanced genetic techniques, de-extinction, or reviving extinct species through techniques like **selective breeding, cloning, and genetic engineering has become a possibility.**
- However, de-extinction projects are **resource-intensive and complicated**, and estimating the probability of success and associated costs is necessary before starting a project.

Methods Used So Far

1. Cloning

- **Uses preserved DNA** to clone extinct species.
- **Example:** Pyrenean Ibex revival (2009).

2. Genome Editing (CRISPR)

- **Modifies DNA** of living species to incorporate extinct genes.
- **Example:** Woolly Mammoth using Asian elephant DNA.

3. Back-breeding

- **Selective breeding** to recreate extinct species' traits.
- **Example:** Aurochs via modern cattle breeding.

4. Synthetic Biology

- **Artificially creates genomes** using DNA synthesis.
- **Still experimental** for species with no close relatives.

Biological Background of Back Breeding

- Back breeding is selective breeding of related species to recreate an extinct species’ **genome**.
- **Genetic Similarity:** Back-breeding uses species with similar DNA to the extinct species to maximize genetic overlap.
- **Mendelian Inheritance:** Offspring inherit genes randomly from parents according to Mendelian inheritance patterns.
- **Genetic Variation:** Genetic diversity is introduced naturally, while breeders select for traits that resemble the extinct species.
- **Mutations:** Mutations occur naturally and can either accelerate or impede the recovery of the extinct species' original genome.

Methodology

- Model Overview
- Simulation Setup
- Simulation Details

Model Overview

Simplified Toy Model for Genetic Recombination:

- Two organisms per generation.
- **Chromosome Setup:** Genes are randomly combined and passed to the next generation.
- **Mutation Factor:** A certain probability of mutation introduces new genes from the gene pool in each generation.

Key Elements:

- Initial organisms: Arrays A1A_1A1 and A1′A_1'A1′ represent the genes (chromosomes) of two organisms.
- Each subsequent generation randomly inherits genes from the previous one.

Goal: Determine how many generations are needed to replicate the exact genotype of either original organism.

Simulation Setup

Master Gene Pool: A fixed set of genes, for simplicity we consider 10 genes in the pool (labeled 1–10).

Chromosome Arrays: Each organism in the first generation has an array of chromosomes, which can repeat genes.

Process:

- Each generation takes random genes from the gene pool and combines them to form new chromosome arrays.
- **Mutation Probability:** A small chance that one gene in the array will mutate (replaced by a new gene from the pool).

Simulation Variables:

- Chromosome count (m= 3, 4, 5)
- Mutation probability (ppp) from 0.01% to 100%.

Stopping Condition: Simulation stops when an organism matches the genotype of an original ancestor.

Simulation Details

- Create initial gene arrays A_1 and A_1'.
- For each generation, genes are randomly recombined to form new arrays A_n and A_n'.
- Apply mutations based on probability p.
- Repeat until a match is found between an organism and the original genotype.

Eg. Gene pool: {1, 2, 3,...,10}

Organisms in first generation- {1, 2, 3, 3}, {2, 3, 4, 9}

After first round of recombination- {1, 3, 3, 4}, {2, 2, 3, 9}

If mutation probability is 30%, there's a 30% chance that one of the genes changes to something else- for example the first organism might now become {1, 8, 3, 4)

Results

m = 3

- As mutation probability increased from 0.01% to 0.2%, there was a sharp increase in the number of generations needed.

Mutation Probability	Number of Generations
0.01	1831063.898
0.05	77583.598
0.1	18935.59
0.15	11894.524
0.2	6859.836

m = 3

After p = 0.2, increasing the probability p didn't lead to as sharp a drop, in fact it was almost linear.

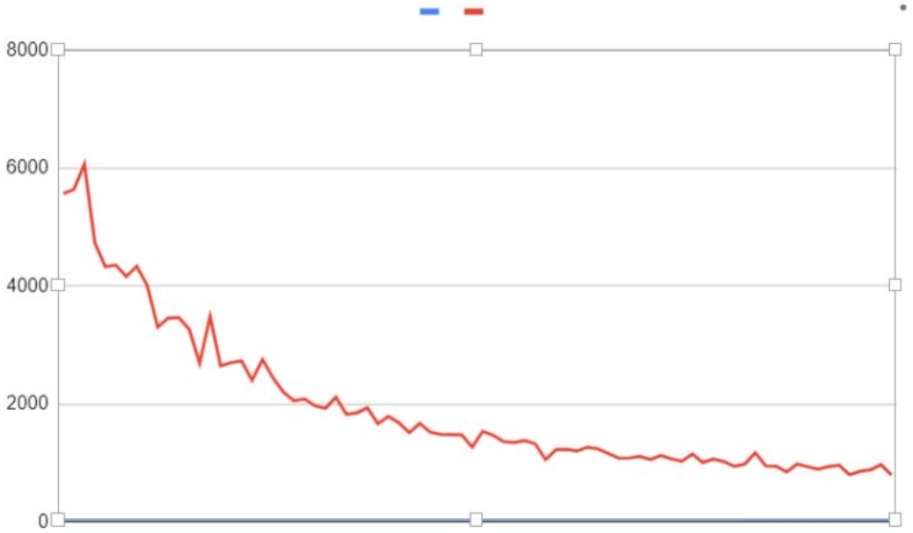


Figure 1: The variation in number of generations as probability of mutation ranges from 0.2% to 1% for m = 3

m = 3

As the probability ranged from 1% to 100%, the number of generations needed was almost constant.

Figure 2: The variation in number of generations as probability of mutation ranges from 1% to 100% for m = 3

m = 4 and m = 5

For m = 4 and m = 5, simulations for p ranging from 0.01% to 1% couldn’t be carried out due to computational constraints. As the probability ranged from 1% to 100%, the results were similar to for m = 3 with no significant trend.

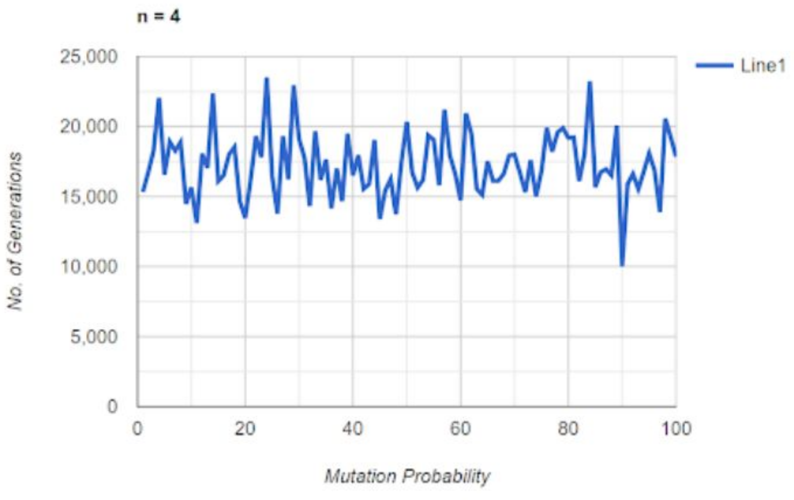


Figure 3: The variation in number of generations as probability of mutation ranges from 1% to 100% for m = 4

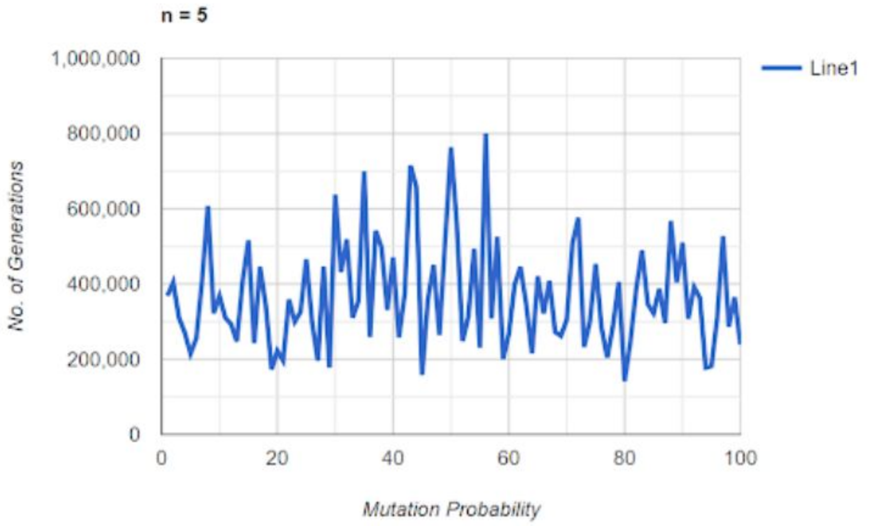


Figure 5: The variation in number of generations as probability of mutation ranges from 1% to 100% for m = 5

Statistical Observations

While there was no clear trend in number of generations n as p ranged from 1% to 100%, the statistical distribution of n for p in the set {1, 2, 3,...,100} is interesting.

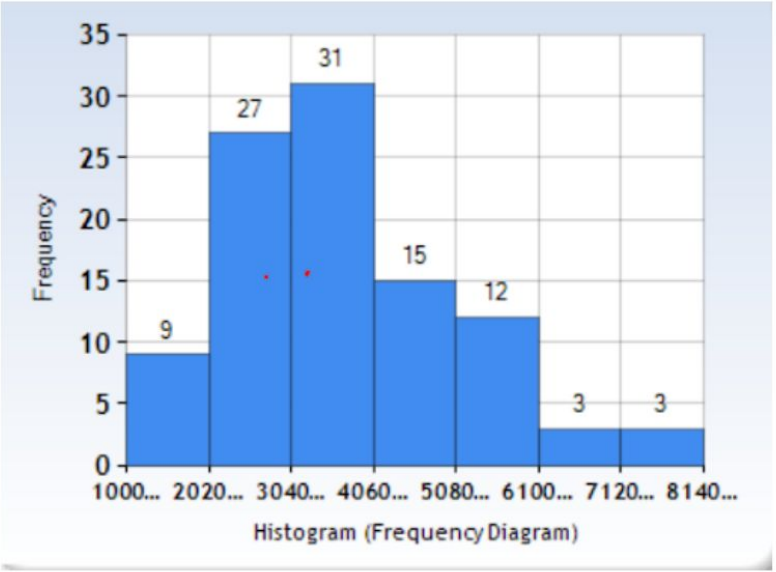


Figure 6: Histogram showing distribution of values of n for p ∈ (1%, 2%,..,100%) for m = 5

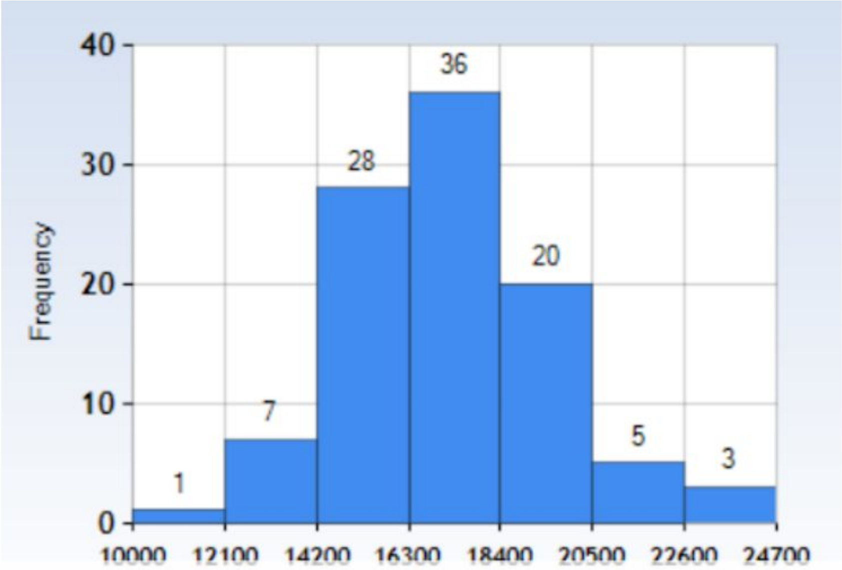


Figure 4: Histogram showing distribution of values of n for p ∈ (1%, 2%,..,100%) for m = 4


Observations and Discussion


Observations

Significant Drop in Generations for m = 3:

- At low mutation rates, small changes in ppp drastically reduce the number of generations.
- For higher p, results stabilize, showing that mutations no longer contribute significantly to genetic recombination at high frequencies.

Behavior for m = 4 and m = 5:

- Similar trends as m = 3, though absolute generations required are higher.

Interesting Statistical Observations:

- Generations follow a Gaussian distribution for $p \geq 1\%$, indicating random spread across possible outcomes.

Conclusions

Key Conclusions:

- Mutation rates between 0.01% and 1% significantly impact the number of generations needed to reproduce an ancestor's genotype.
- After 1%, the required number of generations stabilizes.
- This study provides insights into genetic models for de-extinction and genetic diversity.

Applications:

- Could inform breeding programs for endangered species.
- Helps us understand the role of mutation in genetic conservation.

References

- Novak, B. J. (2018). De-extinction. *Genes*, *9*(11), 548.
- Shapiro, B. (2017). Pathways to de-extinction: how close can we get to resurrection of an extinct species?. *Functional Ecology*, *31*(5), 996-1002.
- Bürger, R. (2011). Some mathematical models in evolutionary genetics. *The Mathematics of Darwin's Legacy*, 67-89.
- Thompson, E. A. (1990). RA Fisher's contributions to genetical statistics. *Biometrics*, 905–914.
- Beskorovajni, R., Jovanović, R., Pezo, L., Popović, N., Tolimir, N., Mihajlović, L., & Šurlan-Momirović, G. (2022). Mathematical modeling for genomic selection in Serbian dairy cattle. *Genetika*, *53*(3), 1105–1115.

Thank You